

AMENDMENT TO THE TITLE

Please amend the title everywhere, the executed declaration excepted, to read -- SYSTEM
AND METHOD FOR EXTRACTING INFORMATION FROM TEXT USING TEXT ANNOTATION AND FACT
EXTRACTION--.

AMENDMENTS TO THE SPECIFICATION

Please replace paragraph [00099] with the following amended paragraph:

[00099] Text annotation is accomplished by individual processes called “Annotators” that are controlled by FEX according to a user-defined “Annotation Configuration.” FEX annotations are of three basic types. Expressed in terms of regular expressions, these are as follows: (1) token attributes, which have a one-per-base-token alignment, where for the attribute type represented, there is an attempt to assign an attribute value to each base token; (2) constituent attributes assigned yes-no values to patterns of base tokens, where the entire pattern is considered to be a single constituent with respect to some annotation value; and (3) links, which connect coreferring constituents such as names, their variants, and pronouns. FEX annotations are captured in a single view of the document expressed as inline XML. In an XML representation, token attributes tend to be represented as XML attributes on base tokens, and constituent attributes and links tend to be represented as XML elements. Shifts tend to be represented as XPath expressions that utilize token attributes, constituent attributes, and links.

Please replace paragraph [00115] with the following amended paragraph:

[000115] Text structure is usually defined or controlled by some type of markup language. In the FEX tool set, an annotated text is represented ~~using as a single view of a document expressed as inline XML,~~ the Extensible Markup Language. The FEX tool set includes a tag uncrossing process to resolve conflicting (crossed) annotation boundaries in an annotated text to produce well-formed XML from the results of the text annotation process prior to fact extraction.

Please replace paragraph [000175] with the following amended paragraph:

[000175] The RuBIE pattern recognition language is a pattern recognition language that applies to text that has been tokenized into its base tokens – words, numbers, punctuation symbols, formatting information, etc. – and annotated with a number of attributes that indicate the form, function, and semantic role of individual tokens, patterns of tokens, and related tokens. Text structure is usually defined or controlled by some type of markup language; that is the RuBIE pattern recognition language applies to one or more sets of annotations that have been captured in a single view of the document and aligned with a piece of tokenized text.

Please replace paragraph [000186] with the following amended paragraph:

[000186] Similar to prior art regular expression-based pattern recognition tools like SRA's NetOwl® technology, Perl™, and the Penn Tools, the RuBIE pattern recognition language supports common, regular expression-based functionality. However, the results of more sophisticated linguistics processes that annotate a text with syntactic attributes are best represented using a tree-based representation. XML has emerged as a popular standard for creating a representation of a text that captures its structure. As noted above, the FEX tool set uses XML as a basis for annotating text with numerous attributes, including linguistic structure and other linguistic attributes, combining these in a single view of the annotated document.